

University of Groningen

Mutual intelligibility in the Slavic language area

Golubovic, Jelena

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2016

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Golubovic, J. (2016). *Mutual intelligibility in the Slavic language area*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 8: General discussion and conclusions

In this thesis we have developed and applied a commensurate measure of intelligibility and factors influencing it. The studies described in the first part of the thesis were a part of a larger project involving Germanic, Romance and Slavic language families. The experiment in the MICReLa web application has been done by more than 40,000 people, over 12,000 of which spoke a Slavic language as their native.

The thesis expands on the work done in the West and South Slavic language area in several important ways. The level of intelligibility between Czech, Slovak, Polish, Croatian, Slovenian and Bulgarian has been measured for the first time in a systematic and comparable way. The distances among those six Slavic languages have been established on five levels: lexicon, orthography, phonology, morphology and syntax. The regression model using linguistic and extra-linguistic factors as predictors can explain up to 66% of variance of the individual intelligibility scores. A teaching intervention aimed at speakers of related Slavic languages has been developed and proved successful in the case of Czech speakers learning to understand Croatian better. Language attitude intricacies between the speakers of Serbian and Croatian, which might have an effect on the intelligibility of the two languages in the future, have been analyzed.

The first two chapters were partly descriptive in nature: we were interested in establishing the level of intelligibility among Slavic languages, as well as the linguistic distance among them at the level of lexicon, phonology, morphology, orthography and syntax. In Chapter 2, we showed that the genealogical distinction into West and South Slavic languages is clear across all linguistic levels, except morphology and syntax, where Bulgarian, with its lack of nominal case and the infinitive form represents an outlier compared to both the West Slavic and the South Slavic group. At this point, we would like to address a limitation in our distance measurements: using the trigram approach, which relies primarily on word order, may not be entirely suitable for a language family exhibiting a high degree of freedom in constituent order such as the Slavic language family. The translators of the material were instructed to stick to the English original as much as possible, deviating from it only if the structure of the language in question required it, or if the resulting construction sounded too unnatural. This means that in all six Slavic languages we received material which corresponded to an outside reference point as much as possible, but it also means that our results might not be generalizable, since the trigram approach cannot capture all the possible orders of

the constituents which are acceptable in a language and cross-compare them. Syntactic distance behaved slightly differently than the other linguistic measurements, but it was still a significant predictor of intelligibility in our dataset. It seems that the syntax measurement, while not ideal, still captures general patterns which explain some of the unique variance in the regression analysis.

In Chapter 3 we measured intelligibility among six Slavic languages using three methods: the word translation task, the cloze test and the picture task. The results of the picture task proved to be unreliable, but the cloze test and the word translation task produced similar patterns of results. This pattern was in line with the genealogical relationships among Slavic languages and the intelligibility results ranged from quite low (for instance for Croatian and Polish) to extremely high, in case of Czech and Slovak, which was quite expectedly the language pair with the highest level of mutual intelligibility. We believe the results are stable and replicable and that the word translation task and the cloze test are suitable tools for assessing intelligibility for large participant groups.

Our results show that the level of intelligibility among the languages investigated ranges from very low (Croatian speakers scored only 9.52% on the spoken cloze test in Polish) to almost native-like (Slovak speakers scored 95.04% on the same task in Czech). The intelligibility of words was, unsurprisingly, higher than the intelligibility of texts, but on the whole, across all methods Czech and Slovak was the most intelligible language combination. The next one in line, Croatian-Slovenian revealed a relatively high level of asymmetry – Slovenian speakers were better at understanding Croatian than vice versa both in the written and in the spoken mode. This asymmetry and the reasons behind it might be an interesting topic of future research. On one hand, it could be the linguistic factors, such as lexicon or syntax, but on the other hand, the level of exposure might also play a role. Young Slovenian speakers do not learn (Serbo-)Croatian at school anymore, but they could still be exposed to Croatian in other ways, such as music, movies, social media, etc. Czech and Slovak speakers can also understand Polish moderately well and the same holds for Polish speakers understanding of Czech and Slovak, while in all the other language combinations we found a relatively low level of intelligibility. If we focus on the results of the cloze test, 24 out of the 30 language combinations we looked at show an intelligibility level lower than 45% in the written mode, while 18 out of those 30 combinations had an intelligibility level lower than 20%. This level of mutual intelligibility is probably not enough for successful communication using receptive multilingualism, except for the very basic communication. At this point we would like to emphasize that it is impossible to place a threshold after which receptive multilingualism should be possible, but it seems safe to assume that less than 20% of intelligibility in the spoken mode would not be enough to maintain an actual conversation using receptive multilingualism.

In Chapter 4 we looked at our data in terms of groups of speakers of one language reading or listening to another language, the question being: how well can we predict the performance of e.g. native speakers of Croatian as a group reading or listening to Slovak? In this case, we only used linguistic factors as predictors in a regression analysis and showed that this way we could account for up to 82% of the variance in intelligibility results.

The linguistic factors used to predict intelligibility in Chapter 4 vary only on the language level i.e. the lexical distance in our dataset is the same for all speakers of Czech and Polish. But extra-linguistic factors, language attitude and language exposure, and of course, our criterion variable, intelligibility, were measured for each participant separately. For this reason, in the model of intelligibility with both linguistic and extra-linguistic factors, presented in Chapter 5 we conducted the regression analysis on the participant level. This inevitably resulted in a smaller percentage of explained variance, but even so, our models could account for up to 66% of the total variance. The predictors we worked with were designed with groups of speakers in mind, but should we wish to predict intelligibility of individuals to an even greater degree, factors such as their working memory, general intelligence or language aptitude should be taken into account. For instance, Vanhove (2014) did an experiment with native speakers of German doing a cognate-guessing task in Swedish. He found that the score on the German vocabulary test (WST) which is a measure of crystallized intelligence and Raven score, a measure of fluid intelligence, were important predictors of the score on the cognate-guessing task, whereby fluid intelligence was a more important predictor in the spoken modality and crystallized intelligence was more important for the score in the written modality.

Another possible limitation of our studies concerns the use of our actual testing material to measure linguistic distances among languages. If intelligibility was measured on the basis of the same material as linguistic distances, the question is how generalizable our results really are. We used two different datasets for both measuring distances and testing intelligibility: the list of 100 words (list data) and four texts of about 800 words in total (text data). The distance results obtained on the basis of the list data were used for predicting the intelligibility results measured with the word translation task, while the distance results obtained on the basis of the text data were used for predicting the intelligibility results measured with the cloze test. There was only one exception: we measured phonological distances on the basis of the list data only and used that as a predictor of intelligibility in both tasks. In both cases, phonological distances were the best predictor of intelligibility. In addition, the results of lexical and orthographic distances measured on the basis of the list data and those measured on the basis of the text data were correlated very highly, both sets of measurement correlate significantly with the intelligibility results and using the list data vs.

the text data results makes almost no difference in the final models. These findings indicate that the distance patterns are replicable and that it is not so important which dataset is used, provided that it is large enough to enable stable measurements.

A measurement that would need improvement is that of language attitudes. Rating how beautiful a language is may well give us a good indication of the attitude towards the language in question, but that might not be readily translated into a factor influencing intelligibility. What might have a greater role as a predictor of intelligibility is the motivation to understand a related language and hopefully future studies will come up with a way to measure that variable effectively. Nevertheless, even motivation (or the lack thereof) to understand a language might not have much of an impact on the participants' performance on a task, since they might be led by a stronger motivation to perform well in the test situation.

Based on the intelligibility results, we concluded that the potential for using receptive multilingualism in the Slavic languages area is limited to a few language combinations. For this reason, we decided to test whether the level of intelligibility could be increased with a very short teaching intervention. The study described in Chapter 6 is different from the rest not only because of its didactic perspective, but also because here we opted to test intelligibility by observing actual interaction between two interlocutors, each of whom was speaking his/her native language. We found that the teaching intervention increased the performance of the experimental group on the intelligibility task by about 105%, while the control group improved only about 27%. We quantified the success of the communication as the number of differences found in a spot-the-difference task, but a more qualitative approach might reveal more about the nature of the RM communication. What strategies do the interlocutors use to make themselves understood and to understand the other person better? Does a teaching intervention targeted at increasing intelligibility have an effect on the success of linguistic alignment in receptive multilingualism? Do the meta-communicative devices the interlocutors used change after the teaching intervention? Bahtina (2013) analyzed the data from chats involving Russian and Estonian native speakers and showed that the dialogues contain examples of alignment at the conceptual, lexical and syntactic levels. Analyzing RM dialogues in a more qualitative way could reveal more about the nature of RM communication and the potential effect the instruction might have on it.

The teaching instruction itself could also be developed more. We made the very first step in discovering that only 4.5 hours is enough to make a difference in intelligibility, but one of the next steps might be creating a longer and more comprehensive teaching program. In addition, the success of teaching interventions could be tested on language combinations with a very low base level of

intelligibility. On one hand, this means that the instruction would have to be more detailed than the one we used here, simply because there is much more to learn. On the other hand, if the base level of intelligibility is very low, it is possible that with the right approach, it could be increased dramatically. This approach would probably have to be quite comprehensive, but the basic grammatical structure, which is fairly similar across most Slavic languages would be taught in terms of correspondences.

Chapter 7 revealed that speakers of Serbian and Croatian hold negative attitudes to each others' languages, which is particularly pronounced in the case of Serbian young teenagers. Interestingly enough, the attitudes of adults who do remember the war years were not so negative, probably due to greater exposure to both standards in their childhoods. Younger Serbian speakers nowadays perceive Croatian as more exotic than their parents do and if their answers to the open ended questions in the questionnaire are anything to go by, they see it as a strange version of Serbian elevated to the status of a separate language. If these findings indicate a trend, i.e., that younger generations tend to be more negative to the neighboring language than older ones, this might have an impact on the intelligibility of Serbian and Croatian in the long run. At the moment, the two languages are so similar than it is difficult to imagine difficulties in communication. Still, the words for new concepts are now coined separately, resulting in a larger number of non-cognates. If we pair that with the negative attitudes, which might affect the motivation for mutual understanding, it does not seem all that unlikely that the intelligibility between Serbian and Croatian might drop in the next few decades.

One of the motivations for our research was finding out where a relatively high level of intelligibility was available with practically no additional training. In the Slavic language area, the only language pair definitely fitting that description is Czech and Slovak. Meaningful communication might be carried out between the speakers of Croatian and Slovene as well, but due to the high level of asymmetry, there would probably be more problems on the Croatian speakers' side. Many language combinations in the Slavic language area exhibit such a low level of mutual intelligibility that speakers of e.g. Slovene would have to learn Bulgarian as a completely foreign language. Of course, they would notice cognates and similar syntactic constructions, but much of the information would be relatively new. By contrast, in language pairs with a moderate degree of intelligibility, such as Croatian and Czech, a targeted instruction such as the one we described in Chapter 6, could increase intelligibility level significantly in a relatively short time. This is not to say that the difference between language pairs with high, medium and low level of intelligibility is definite and clear-cut. Rather, for different language pairs the same communication and learning strategies might be successful to a different extent, depending on their level of intelligibility.

Europe is the home to more than 200 indigenous languages.²⁵ We see receptive multilingualism as a way to maintain this rich linguistic diversity within the same language families, while not compromising the effectiveness of communication. Of course, this is not possible for all the language combinations equally, but as we have showed, a little instruction can go a long way. The greatest problem with using receptive multilingualism is probably the general lack of awareness about it. Except the language combinations such as Czech and Slovak, where it has been used for decades, Slavic speakers generally do not use receptive multilingualism and do not have too much exposure to the speakers of other Slavic languages for that matter. Hopefully raising awareness about receptive multilingualism as a viable option will make the Slavic language area a much smaller and better connected place.

25 <http://www.ethnologue.com/statistics/area#1> retrieved on 12.08.2015.